

Edge Detection With Direction Guided Postprocessing for Farmland Parcel Extraction

Yusen Xie , Shaolan Zheng, Haiyun Wang, Yuzhou Qiu, Xilan Lin, and Qian Shi , *Senior Member, IEEE*

Abstract—Farmland is a significant resource for human survival and development. Rapid acquisition of farmland information is the basis for dynamic crop detection and sustainable land development. The continuous development of high-resolution remote sensing imagery makes it possible to make a wide range of refined earth observation. With better image interpretation ability, image segmentation method based on deep learning can bring specific results from high-resolution imagery and is widely used in remote sensing. However, existing image segmentation methods based on semantic segmentation have difficulties to extracting refined farmland parcels. Deep neural network is used to detect farmland edge. We use high-resolution network to achieve feature extraction that retains high-resolution features, strengthens the feature representation of network context information based on object-contextual representations module, and carries out more complete interpretation of farmland and its boundary. Finally, we design a farmland edge postprocessing method to connect the disconnected boundary based on the direction information generated by the connectivity attention module, and finally obtained the farmland boundary which is complete enough to be closed for generating farmland parcels. To verify our method, we used Google Earth image to label farmland boundaries and conduct experimental verification. The results show that our proposed model has a higher precision for farmland edge detection, and the postprocessing method of boundary connection can effectively close the boundary lines and achieve more detailed and complete farmland parcels.

Index Terms—Deep learning, edge detection, farmland extraction, fully convolutional networks, semantic segmentation.

I. INTRODUCTION

HUMAN survival relies on agriculture, and farmland is one of the bases for agricultural production [1]. Accurate information extraction of farmland areas has become an urgent

requirement for precision agriculture and sustainable development [2]. Precision agriculture requires high-efficiency and low-cost benefits to be obtained in an efficient and economical manner under the support of advanced technology [3]. Farmland is the first unit of agricultural operation and production, so how to efficiently obtain refined farmland parcels is the basis of precision agriculture research.

The use of artificial ground survey methods is costly and inefficient. The development of remote sensing technology provides a data basis for large-scale and efficient farmland parcel research, so the production of farmland parcels through artificial visual interpretation has become the mainstream method for practical application and research [4]. However, this method still fails to solve the problem of inefficiency. Classic image segmentation and edge detection methods, such as watershed algorithms [5] and canny [6], are first tried in farmland extraction [7], [8], [9]. Objected-based image analysis are widely used in automated farmland parcel extraction. Based on the brightness gradient information and texture information of the image, this type of method uses super-pixel methods to segment the remote sensing image into multiple independent and homogeneous objects, and classifies the segmented objects through SVM, random forest, or other machine-learning methods to obtain farmland parcel results [10], [11], [12], [13], [14]. However, with the improvement of remote sensing image resolution, the performance of traditional image segmentation methods is limited in richer ground object texture features.

Deep-learning models are heavily used in different types of image processing tasks, including image classification, object detection, and semantic segmentation [15]. Many classic convolutional neural networks, such as AlexNet, VGGNet [16], ResNet [17], DenseNet [18], and SENet [19], have achieved superior performance in image classification. Full convolutional network [20] proposed a new paradigm of image processing, which provides per pixel predictions, such as SegNet [21], U-net [22], DeepLab series [23], [24], [25], [26], and DANet [27].

Deep-learning models have achieved good results in different types of image processing problems, and scholars have also tried to apply them in remote sensing [28], [29], [30], [31], [32], [33], [34], [35], [36], [37]. Researches have shown that semantic segmentation methods have a better accuracy and robust prediction in high-resolution satellite or aerial imagery [38], [39], [40], [41], [42]. Semantic segmentation models are applied and modified in farmland extraction [43], [44], [45], [46], [47], based on context representations, multiscale or pyramid features, and attention modules. These networks have the benefits

Manuscript received 25 December 2022; revised 8 February 2023; accepted 21 February 2023. Date of publication 7 March 2023; date of current version 20 April 2023. This work was supported in part by the National Key R&D Program of China under Grant 2022YFB3903402, in part by the National Natural Science Foundation of China under Grant 42222106, and in part by the National Natural Science Foundation of China under Grant 61976234. (*Corresponding author: Shaolan Zheng.*)

Yusen Xie and Qian Shi are with the Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: xieys3@mail2.sysu.edu.cn; shixi5@mail.sysu.edu.cn).

Shaolan Zheng, Haiyun Wang, Yuzhou Qiu, and Xilan Lin are with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou 526040, China, also with the Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical Area of South China, Ministry of Natural Resources, Guangzhou 510650, China, and also with the Guangdong Science and Technology Collaborative Innovation Center for Natural Resources, Guangzhou 515063, China (e-mail: shaolan51@163.com; hywang900608@gmail.com; gzhuqyz@163.com; 3155632139@qq.com).

Digital Object Identifier 10.1109/JSTARS.2023.3253779

of enhancing [6] semantic representations by larger receptive fields and deeper networks. Thus, complex contours between objects may be neglected or ambiguous, and this issue becomes more fatal in refined farmland parcels extraction. Details inside farmland are easily lost and finally models can only bring coarse results. Hence, edge detection methods are attempted in farmland parcels extractions, which can focus on the junction area between farmland parcels and bring more details [48], [49]. However, breaking points occurs in the edge results from edge detection model, which makes it difficult to produce vectorial farmland parcels.

In this article, we propose a process based on edge detection to extract the boundary between farmland parcels through the semantic edge detection model based on deep learning, and finally close the boundary into vector parcel results. The proposed semantic edge detection model is based on the high-resolution feature extraction network, and the context information feature representation of the network is enhanced based on the object-contextual representations (OCR) module. In order to improve the extracted boundary that can be closed into a polygon as much as possible, we propose a postprocessing method based on breakpoint connection, which introduces the connectivity attention module (ConAttn) into the edge detection model to generate direction information through supervised training, and connects the breakpoints based on the direction information and edge confidence, which can finally make the extracted boundary results as complete as possible and obtain more accurate farmland parcels.

The rest of this article is organized as follows. Section II demonstrates the proposed network architecture. Section III presents experiments and the corresponding analysis with edge detection models and the proposed postprocessing methods. Section IV discusses superiorities of edge training labels. Finally, Section V concludes this article with some remarks and hints at plausible future research lines.

II. METHODS

In this section, we design an edge detection network based on semantic segmentation structure. The feature extraction part of semantic segmentation structure contains multiple parallel multiscale resolution feature branches, which enable network to learn stronger semantic information and precise position information at the same time, and each branch is composed of multiple residual blocks. The upsample part of semantic segmentation structure adds the feature representation enhanced by the model to the context information, and generates multiscale features that represent both object information and background information through convolution. The information strengthens the extraction of farmland features. Some approaches are designed to enhance the farmland edge. A ConAttn generates direction information, and an automatic boundary line connection method is introduced to close the edge. Finally, closed edge lines are converted into polygon results as farmland parcels.

A. Semantic Edge Detection Architecture

For refined farmland block extraction tasks, the traditional semantic segmentation model is used to train with polygon

labels, which can accurately identify a large range of farmland areas. However, the model trained with polygon labels is not sensitive to the edge between farmland parcels. The adjacent parcels cannot be successfully identified separately, and the refined parcels cannot be obtained. Therefore, we need to use farmland edge as labels, so that we can choose to use semantic edge detection model. Through experiments, it is found that the traditional edge detection model is difficult to converge in the farmland edge detection task. Compared with the edge of all objects in the natural image, the edge detection network focuses on the lower image features, while the farmland edge has stronger semantic information. So the network needs to fully extract the semantic features of the farmland. Therefore, we use the semantic-segmentation-based network as our final farmland edge detection network structure.

Semantic segmentation network structure contains the feature extraction part and the upsample part. The feature extraction part is used to compute image representations by multiple convolutional layers, and obtains deep semantic features when the image resolution decreases with the network structure going deeper. Based on the spatial position information corresponding to the extracted deep semantic feature, the upsample layer restores the resolution of the image representations to be consistent with the input image through the upsampling structure, and finally outputs the classification information corresponding to each pixel through the classifier layer to obtain the pixel-level image classification result. In the edge detection task in this article, the semantic-segmentation-network-based structure extracts farmland semantic information by entering high-resolution remote sensing satellite imagery, and outputs the pixel-level farmland edge results.

B. High-Resolution Network

High-resolution network (HRNet), originally proposed by [50], is a feature extraction network that can maintain high-resolution feature information. The commonly used semantic segmentation network is based on the feature extraction layer in the image classification task as the backbone, so the structure of these backbones usually shrinks the feature image resolution as the network deepens, so that the network can learn deep semantic features. However, in the semantic segmentation task, the network ultimately needs to output pixel-level prediction results, so it also needs to upsample low-resolution feature maps. With the process of downsampling and upsampling to feature maps loses a lot of spatial details, thereby limiting the accuracy of pixel-level classification results. To solve this problem, common semantic segmentation networks, such as SegNet [21] and U-net [22], are introduced into the feature extraction layer through the up-sampling process, but the simple skip connection combined with shallow features lacks feature consistency. HRNet retains feature information at different scales through parallel multiresolution branches, and retains strong semantic information and spatial location information in features through information interaction between different resolution branches, which are appropriate for high-resolution satellite imagery with complicated details.

The overall structure of our proposed network is shown in Fig. 1. The HRNet contains four stages. As the network deepens,

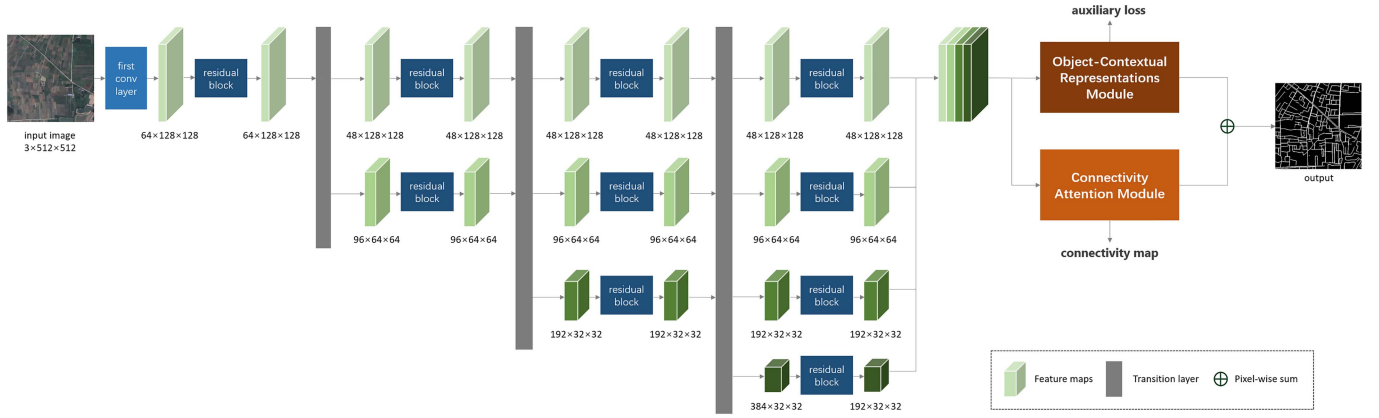


Fig. 1. Overview of the proposed network architecture. The backbone of the network is HRNet, comprised of parallel branches of multiscale resolution features. At the end of the backbone, all feature maps are rescaled to 128×128 and concatenated. Then, objected-contextual module enhances the representation of context information and outputs the farmland edge results. Connectivity attention module is designed to extract information of connection and direction. The module can output connectivity map and optimize edge results by a pixelwise sum fusion.

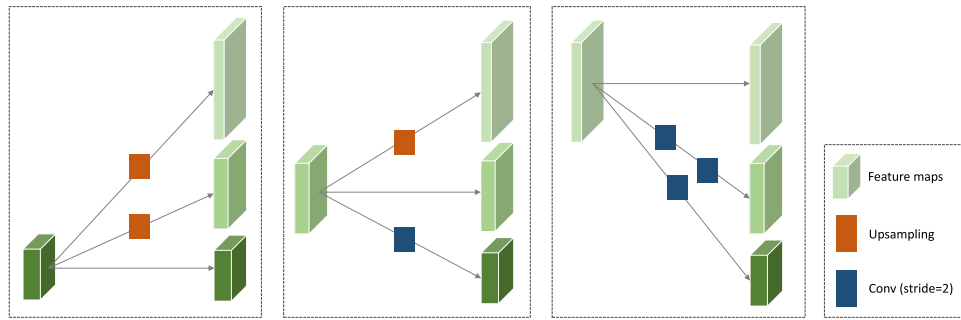


Fig. 2. Transition layer. The upsampling layer is based on bilinear interpolation. Convolution with stride = 2 is used to downsample feature maps.

a lower resolution branch will be added between the stages, and the features between different resolution scales will be interacted with through the transition layer. The resolution of feature N_1 is r_1 , and the resolution of the q layer is $r_q = \frac{1}{2^{q-1}}$. In order to realize the information interaction between multiscale branches, in the feature transition layer, feature is transferred by dense connections between different branches. Taking a feature transition layer with three branches as an example, the features corresponding to three different resolutions in the three branches can be represented as $\{N_1^i, N_2^i, N_3^i\}$. After passing through the transition layer, the branches of input features with the resolution of r_q can be represented as $N_q^{i+1} = f_{1-q}(N_1^i) + f_{2-q}(N_2^i) + f_{3-q}(N_3^i)$, where $f_{x-y}(\cdot)$ denotes a transform module inside the feature transition layer, as shown in Fig. 2. When $x = y$, $f_{x-y}(N_x^i) = N_x^i$; when $x < y$, feature maps need to be downsampled to a lower resolution; when $x > y$, feature maps need to be upsampled to a higher resolution, so $f_{x-y}(\cdot)$ corresponds to bilinear upsampling, which upsamples feature maps to the resolution corresponding to the target branch, and align the number of channels through 1×1 convolution.

C. Object-Contextual Representations

The proposed method needs more semantic information to detect farmland boundaries, and the adjacent farmland parcels are contextually connected. Thus, the upsample module uses the

OCR module, shown in Fig. 3, first presented by [51], which strengthens context aggregation in semantic segmentation. The network structure enhances the representation of context information based on the target area of the object, thereby perfecting the pixel-level classification results with the context information. First, the spatial features extracted by the backbone network output pixel representation through the convolutional layer.

To better integrate object-based semantic information and dense spatial context information, the model refers to the cross-attention structure in the transformer [52]. Soft object regions, which contains object category information, generates weights through Softmax function, and multiplies them by pixel representation. Object region representation is obtained by the sum of pixel-level feature representations and object information. The calculation formula can be expressed as

$$w = \text{Softmax} \left(\phi_1(p)^T \phi_2(o) \right) \quad (1)$$

where p denotes pixel representation, o denotes soft object regions, and $\phi(\cdot)$ denotes transformation function including 1×1 convolution, batch normalization, and ReLU. Object region representation obtains object-based context information by inner product, and implements cross-attention with pixel representation, mapping the object information containing context to a dense feature space to obtain OCR, which can be

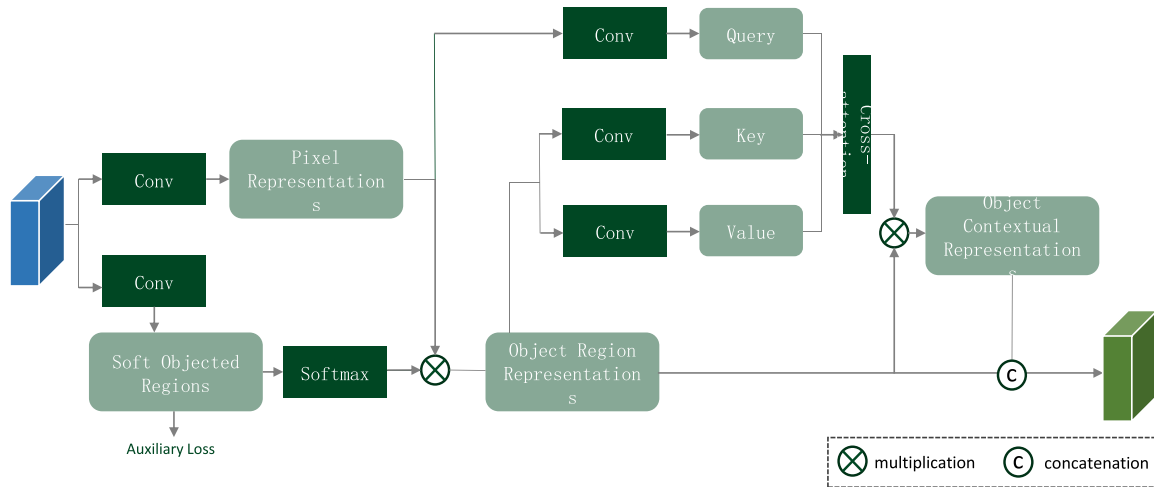


Fig. 3. Object-contextual representations module.

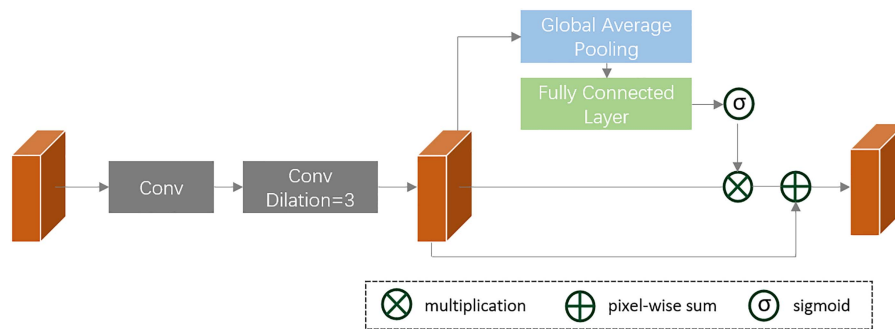


Fig. 4. Connectivity attention module.

expressed as

$$\text{OCR} = \text{Softmax} \left(\phi_3(w)^T \phi_4(w) \right) \phi_5(p). \quad (2)$$

D. Postprocessing

The edge obtained in the edge detection method based on deep learning cannot guarantee the connectivity of boundary lines because it is generated by probability maps. However, the method of generating parcels based on boundary lines needs to ensure that the boundaries are connected and closed into synthetic parcels, so more postprocessing methods need to be introduced. In this article, an edge breakpoint connection method based on direction information is proposed. First, the breakpoint detection of the skeletonized boundary is performed, and all breakpoints are placed on the stack. In each iteration, first, a breakpoint is taken out of the stack. The direction information of the point and the edge probability map in the neighborhood are combined to determine the connection direction. The direction information and boundary probability maps above are obtained by the deep-learning model, and the module that outputs the direction information will be introduced in the next section. When getting the target connection direction, the pixel in the

target direction will be turned into new edge area. A pseudocode of the edge breakpoint connection is shown in Algorithm 1. Finally, if the new boundary pixel is a breakpoint, put the pixel on the stack. Through the above iterative process, the edge detection results can be completely closed, thereby improving the integrity of the farmland parcel results.

E. Connectivity Attention

In the previous section, the introduction of direction information is critical in connecting breakpoints. If only the probability plot is used as a guide, the method connects the low-confidence areas adjacent to the boundary, and the results are similar to the double threshold postprocessing method in Canny [6]. The problems still remains that some breakpoints cannot be connected with others during the connection process. In this article, a connectivity map representing direction information is used, and the ConAttn, shown in Fig. 4, is introduced into the edge detection model to train and predict the connectivity map corresponding to the boundary of the image. Finally, the connectivity map provides direction information to guide the breakpoint connection in the postprocessing process.

Algorithm 1: Breakpoint Connecting Pseudocode, Python-Like.

```

# sk: convert edge probability maps to binary line maps
# bp: breakpoints detection function
# bpd: breakpoint detection for the given point
# dg: direction guided function
# con: extend breakpoints by the given direction
x_line = sk(x)
bp_list = bp(x)
poi_stack = deque(bp_list) # initialize the stack for
breakpoints
while len(poi_dq) > 0:
    poi = poi_stack.pop()
    dr = dg(poi, x)
    x_line, new_poi = con(x_line, dr, threshold = 0.01) #
    threshold to exclude no edge probability area
    if bpd(new_poi, x_line):
        poi_stack.append(new_poi) # push the new
        breakpoint into stack

```

The connectivity map quantifies and disassembles the connection direction of the edge to strengthen the learning ability of the model for the connective direction features.

For parcel boundary labels, connectivity maps $O \in R^{H \times W \times C}$ are generated, where C represents the number of samples of a given pixel and its neighbors. C is denoted as 8 for a 2-D edge image, and $O_{i,j,c}$ in the connectivity map represents the connectivity between the pixel and the specific pixel, where i, j represent the spatial position of the pixel. c represents the location of its adjacent pixels, and if two pixels are connected, it is recorded as a positive label, indicating that the two are connected farmland boundaries. Finally, the connectivity information in the eight neighborhoods is calculated and concatenated to obtain the final connectivity map.

The module used to generate the connectivity map uses the design of the squeeze-excitation [19] structure. The features generated by the backbone network are input into the 3×3 convolution with expansion rates of 1 and 3 in turn, and the use of convolution with different expansion rates can expand the receptive field and fully mine a larger range of local information. Then, the features are pooled globally. The feature is compressed at the channel level, and the feature attention is learned by using two fully connected layers to obtain a vector with a range of (0, 1). The vector can be multiplied by the input feature to obtain the connectivity attention, and finally supervised learning with the connectivity map ground truth generated by the label. The loss function can be expressed as

$$L_d = - \frac{1}{C_O \times N} \sum_{c=1}^{C_O} \sum_{i=1}^N \left(y_i^C \cdot \log(\widehat{y}_i^C) + (1 - y_i^C) \cdot \log(1 - \widehat{y}_i^C) \right) \quad (3)$$

where C_O denotes the number of channels in the connectivity map; $N = H \times W$ means the sum of pixels; y_i^C denotes the



Fig. 5. Samples of the farmland edge dataset with red marks.

ground truth of connectivity map in the location i ; and \widehat{y}_i^C denotes the prediction of connectivity map.

III. EXPERIMENTS AND ANALYSIS

A. Datasets

We yield a farmland edge dataset based on Google Earth high-resolution satellite imagery. Images are uniformly resampled into 0.5-m ground resolution using bilinear interpolation method. First, satellite imagery covering the countryside area in Guangdong province are selected, and we screen out images in mountainous and plain areas. After manual annotation with artificial visual interpretation, we finally got farmland edge dataset containing 2873 tiles and 512×512 images. Edge labels of farmland are converted into binary maps with a width of three pixels as the training labels. Fig. 5 shows experimental images of the datasets. In the subsequent experiments, the datasets are divided into training set, validation set, and test set at a ratio of 6:2:2. The training set and validation set are used for training, and the test set is used for assessment and analysis.

B. Evaluation Metrics

In evaluation part, we use four metrics for pixel-based evaluation, including precision, recall, $F1$ -score, and intersection over union (IoU). These metrics are widely used for accuracy evaluation in semantic segmentation tasks, and in this article, these metrics are used as the criteria in edge detection. $F1$ is the weighted average of precision and recall. IoU calculates the ratio of the intersection and union between the predicted category and the real category. Indicators above are explained as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (7)$$

where TP, FP, and FN denote the true positive, false positive, and false negative, respectively.

C. Implementation Details

Data augmentation strategies are implemented to enhance generalization ability of model. Spatial augmentation methods include random clip and resize, random rotation at different angles (90°, 180°, 270°, and 360°), and random mirror flip vertically or horizontally. Spectral augmentation methods include brightness and contrast enhancement, blurs, and gauss noise.

Network training was implemented by Pytorch using two GeForce RTX 2080Ti. Backbone parameters of all edge detection networks are pretrained on ImageNet [53], and the rest parameters of networks use Kaiming initialization methods [54]. AdamW [55] with the initial learning as 0.0002 is adopted as training optimizer to make the training process converge quickly. Batch size is set to 16. During the training process, the learning rate is adjusted adaptively based on the loss. When the loss stops decreasing for three consecutive epochs, the learning rate will decay with a rate of 0.5, and the training stops when the learning rate is less than 1e-7. We use binary cross-entropy loss as the loss function in experiments.

D. Compared Models

To evaluate our proposed methods in farmland edge detection, some classic semantic segmentation networks, such as DeepLab V3+ and LinkNet, are introduced to make comparison with our proposed model. Meanwhile, D-LinkNet is considered in the comparison for its good performance in the road extraction task from high-resolution satellite imagery, and the task is similar to farmland edge detection in this article. A brief introduction of these models are as follows:

- 1) *DeepLab V3+* [26]: Optimized from a series of Deeplab networks, the network uses Xception [53] as its backbone. Atrous spatial pyramid pooling module is proposed in the network to encode multi-scale spatial contextual information with convolution kernels at multiple rates. spatial. To this extent, Deeplab V3+ achieved state-of-the-art performance on the PASCAL VOC 2012 semantic image segmentation benchmark [56].
- 2) *LinkNet* [57]: LinkNet brings the residual connection into the most typical encoder–decoder symmetrical architecture model, U-net [22]. When feature maps get upsampled, a cascading contraction path brings corresponding shallow features to the decoder. Network architecture in LinkNet is modified, which uses less parameters but gives state-of-the-art performance on CamVid [58] and comparable results on Cityscapes dataset [59].
- 3) *D-LinkNet* [60]: Built based on LinkNet for its efficiency in computation and memory. Dilation convolution is introduced to enlarge the receptive field of feature points without reducing the resolution of feature maps, so that it can perform better in satellite imagery. The network

TABLE I
QUANTITATIVE RESULTS FOR IOU, RECALL, PRECISION, AND F1-SCORE OF
EDGE DETECTION MODEL COMPARISON

Method	IoU	Recall	Precision	F1
DeepLab V3+	0.6258	0.7654	0.7742	0.7698
LinkNet	0.5203	0.7901	0.5789	0.6845
D-LinkNet	0.5558	0.7669	0.6621	0.7145
Ours	0.7398	0.8346	0.8664	0.8505

gets best IoU scores in the CVPR DeepGlobe 2018 Road Extraction Challenge [61].

E. Experiments and Analysis

In this part, we will first compare the proposed farmland edge detection model with other typical semantic segmentation and road extraction models, and then present the results after postprocessing. The effectiveness in the network module is then demonstrated by ablation experiments.

1) *Edge Detection Model Comparison*: Table I shows the accuracy evaluation part, including IoU, recall, precision, and F1 score. The proposed model is optimal in all indicators, IoU and F1 score obtaining 0.7398 and 0.8505, respectively.

Fig. 6 shows the prediction visualization results of comparative models. The proposed model has better recognition ability of farmland edge in different scenarios. As a traditional semantic segmentation network, DeepLab V3+ extracts more incoherent parts of the edge extracted in the face of boundary extraction tasks, which also exists in our proposed method, which is also based on a traditional semantic segmentation network. However, our model can effectively identify more areas of edge, which is consistent with our relatively high price of IoU indicators in accuracy evaluation. LinkNet and D-LinkNet use the skip connection design in the U-Net structure and use transpose convolution in the decoder to increase the feature resolution gradually, so the overall coherence of the results from LinkNet and D-LinkNet is stronger. However, the overall edge detection accuracy is relatively lower, and there are more missed extractions for the farmland boundary in more details. Since the proposed model enhances the extraction ability of contextual information, the interpretation of farmland areas in the image is stronger, so that more accurate boundary extraction can be obtained. For large farmland parcels, where other models cannot fully identify the boundaries of a single parcel, the proposed model is able to extract the complete edge of large parcels. For densely distributed small parcels, other models tend to lose more detailed edge of complex small parcels, while our model has richer and more complete details. However, it is worth noting that although the accuracy of the boundaries extracted by our model is higher, the number of breakpoints is significantly higher than that of LinkNet and D-LinkNet, so the postprocessing process needs to be introduced.

2) *Postprocessing Results*: After obtaining the farmland edge extracted by the model, we evaluate the postprocessing effect after extracting the edge centerline using the skeletonizing



Fig. 6. Visualization of the edge detection results from comparative models.

method. Comparisons include results from no postprocessing at all, double threshold methods proposed in canny [6], and breakpoint connection postprocessing methods without direction information guide. Fig. 7 displays that there are a large number of disconnected edge lines having been connected after postprocessing. Most of these connection areas are with weak edge features. Deep-learning models often give lower probability values because their features are not obvious enough compared to strong edge features, and the edge obtained after final binarization have a large number of incoherences. The proposed connecting postprocessing strategy focuses on supplementing the areas with low confidence to close more boundaries. It can be seen from the final generated parcel vector results that since more boundaries can be closed into polygons, more parcels can be generated after vectorization. When there is no postprocessing method to connect the breakpoint, even if the boundary of the parcel can be mostly identified and the above accuracy evaluation performance is normal, parcels from the process of

generating vector will still cause a large number of missed extraction farmland parcel. When the postprocessing method can connect more breakpoints, the resulting parcel vector result is more refined.

To numerically evaluate the postprocessing effect, we rethicken the postprocessing boundary, corresponding to the label with the boundary line width of three pixels used in model training, and evaluate the accuracy of the label based on the bold result. Table II shows that the postprocessing of breakpoint connection can effectively improve the accuracy of boundary extraction, and the accuracy can be further improved after introducing direction information. The postprocessing process can also complete the breakpoint connection as much as possible for areas with low confidence in the output boundary from the network. It can also be seen from Fig. 7 that after introducing direction information, the problem that the boundary cannot be connected after the breakpoint is extended can be effectively avoided, and it can be effectively connected to another boundary.

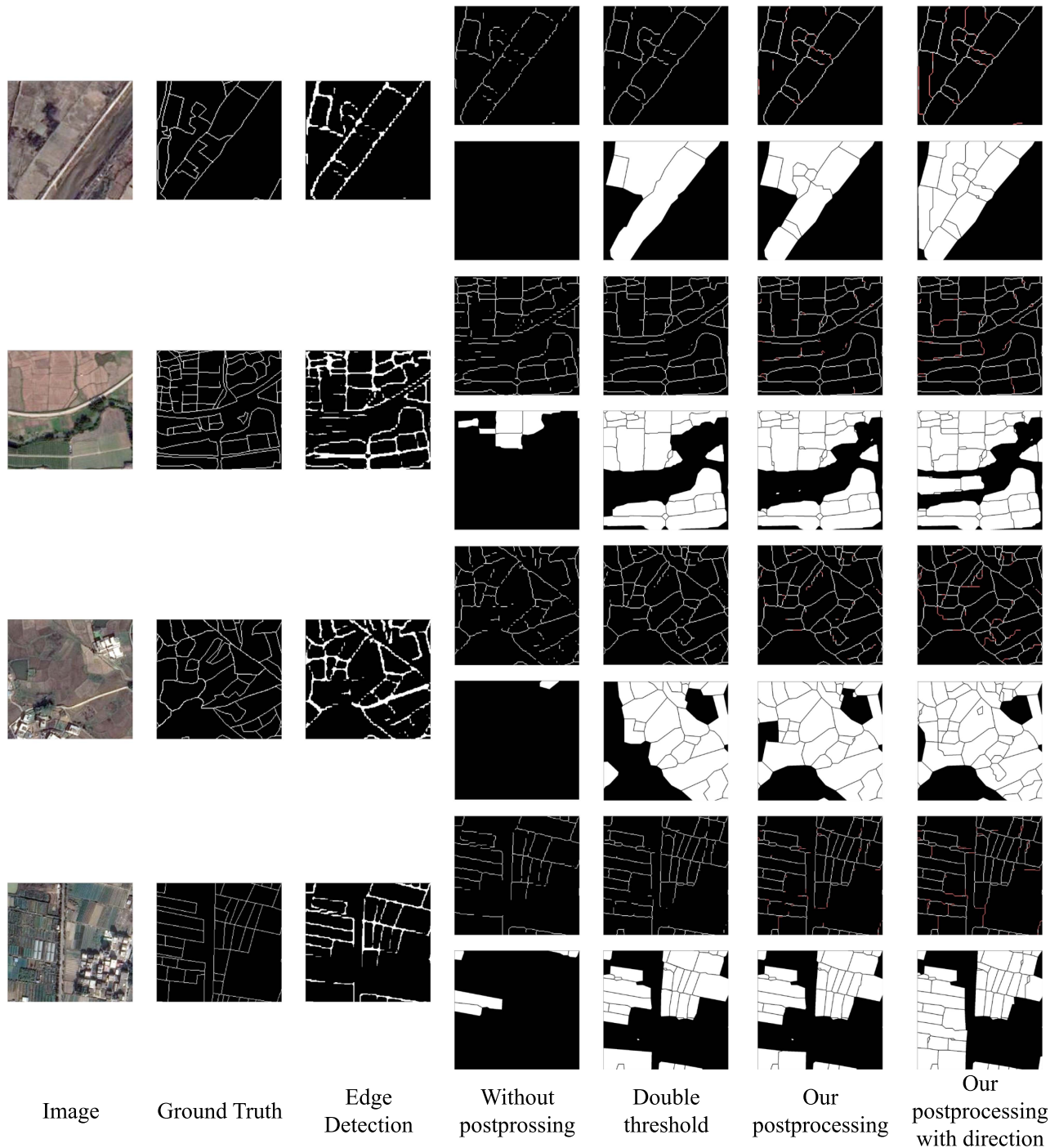


Fig. 7. Visualization of postprocessing results. The polygon results are vectorized by closing the edge line above.

3) *Ablation Experiments*: To demonstrate our effectiveness of the ConAttn added to the edge detection model, we performed ablation experiments on this module in our farmland edge dataset. The results are shown in Table III. The baseline is object-contextual representations network (OCRNet) in our proposed methods. ConAttn without fusion denotes that the ConAttn is added and the training is supervised, but the results output by the ConAttn are not fused with the edge prediction probability maps. After the model introduces direction information from supervised training of the ConAttn, the model

can enhance the sensitivity of the edge and the surrounding connected boundary. There is a certain improvement in various indicators. When we add all the direction information of the connectivity attention result and fuse it with the edge probability map to enhance the edge detection results, the accuracy evaluation indicators have been further improved, indicating that the ConAttn can provide effective direction and connection information. Thus, through the simple probability map fusion, the advantage of perfect boundaries can be exerted.

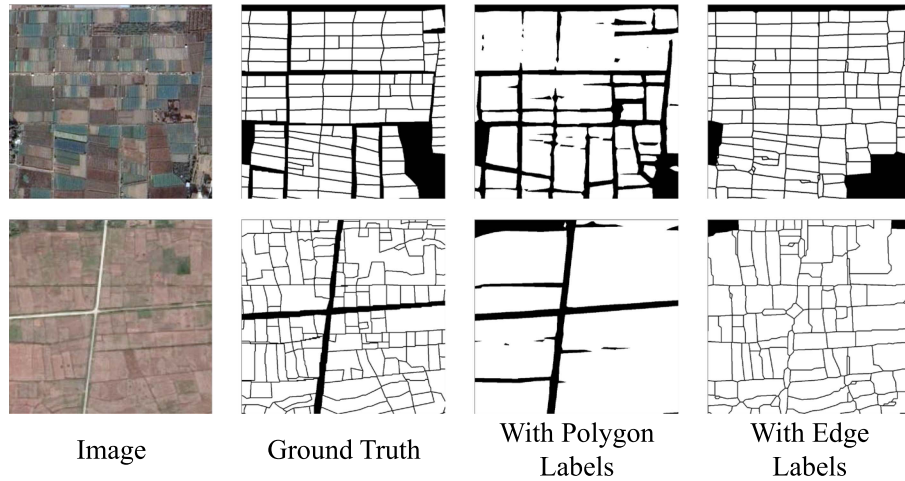


Fig. 8. Comparison with models trained on polygon labels.

TABLE II
QUANTITATIVE RESULTS FOR IOU, RECALL, PRECISION, AND *F1*-SCORE OF
POSTPROCESSING COMPARISON

Method	IoU	Recall	Precision	F1
Without postprocessing	0.5219	0.6317	0.7621	0.6969
Double threshold	0.6071	0.7384	0.8201	0.7792
Our postprocessing	0.6383	0.7808	0.8342	0.8075
Our postprocessing with direction	0.6421	0.8021	0.8524	0.8272

TABLE III
QUANTITATIVE RESULTS FOR IOU, RECALL, PRECISION, AND *F1*-SCORE OF
ABLATION EXPERIMENTS

Method	IoU	Recall	Precision	F1
Baseline	0.6950	0.7968	0.8434	0.8201
Baseline + ConAttn(without fusion)	0.7005	0.8018	0.8460	0.8239
Baseline + ConAttn	0.7398	0.8346	0.8664	0.8505

IV. DISCUSSION

In the discussion part, we compare the proposed methods trained on different types of farmland parcel labels, including edge labels and polygon labels, to discuss possible ways for improvement.

In the method introduction, we mentioned that semantic segmentation methods trained by traditional polygon labels are not suitable for more refined farmland parcel extraction, so we further verify the proposed farmland parcel extraction process based on edge detection, and we compare the extracted results with the output of the semantic segmentation model based on polygon label training. In this part, a same network based on OCRNet is used, and the network is trained by edge labels and polygon labels from the same dataset, respectively.

Fig. 8 is a visualization of the results. After vectorization, the model trained based on polygon labels has higher recognition accuracy for farmland areas. The extracted farmland areas are more accurate, and the model trained based on polygon labels can effectively distinguish from farmland for the thick boundary between farmlands, such as roads. However, it is clear that there are a large number of missing boundary details in the internal parcel, and a large number of parcels with obvious boundaries are aggregated together, so these methods are hard to generate refined farmland parcels. Since the polygon label contains the spatial texture information of the farmland area, and although the boundary area is easily diluted by other features as the supervision signal brought by the negative sample label. The network cannot pay enough attention to the edge area of the farmland, and the farmland texture information is fused together with the decrease of feature resolution during the downsampling process of the feature maps. In the farmland edge detection process we use, since the supervision signal is the edge, the network can focus on the boundary features between the parcels and output the edge containing more detailed information. The postprocessing process we use greatly compensates for the missed farmland caused by the edge failure to be closed into a polygon, so that the edge-detection-based results will decrease missing parcels. However, through this experiment, it can also be seen that the semantic segmentation method based on polygon labels still has advantages for farmland range extraction. For the thick boundary between farmlands, the edge-detection-based method cannot be well divided, and the thick boundary areas are easy to be misclassified into the farmland range. How to combine our processes with segmentation methods trained by polygon labels to form more accurate farmland parcel results is a direction that needs further consideration and optimization.

V. CONCLUSION

In this article, we try to extract farmland edge and generate more refined farmland parcels based on the farmland edge detection results. Based on this task, we propose an edge detection

model based on connectivity attention, which can retain multi-scale image features through high-resolution structure network, and integrate the object-based semantic information and the dense spatial context information through OCR module. The ConAttn improves the model's perception of boundary directions, while outputting direction information for postprocessing. The results show that although the proposed model can extract the edge area of farmland more accurately, there are still more breakpoints, so we propose a postprocessing method based on breakpoint connection, which connects the incoherent locations of the boundaries extracted by the model based on confidence and direction information, and finally allows more edge to be closed into blocks, thereby completing more refined farmland parcels extraction.

In the experiments in this article, most of the farmland comes from plain areas, so in future article, we will continue to improve our methods, and explore the farmland extraction ability of the model in different regions to improve the generalization and transferability of the model. At the same time, we will optimize the efficiency of postprocessing so that it can be used for a wide range of farmland mapping.

REFERENCES

- [1] B. F. Wu, J. H. Meng, and Q. Z. Li, "Review of overseas crop monitoring systems with remote sensing," *Adv. Earth Sci.*, vol. 25, pp. 1003–1012, 2010.
- [2] P. S. Thenkabil, "Global croplands and their importance for water and food security in the twenty-first century: Towards an ever green revolution that combines a second green revolution with a blue revolution," *Remote Sens.*, vol. 2, pp. 2305–2312, 2010.
- [3] Z. Wu et al., "Research on precision agricultural based on the spatial-temporal remote sensing collaboration," *J. Geo-Inf. Sci.*, vol. 22, pp. 731–742, 2020.
- [4] K. A. Addo, "Urban and peri-urban agriculture in developing countries studied using remote sensing and in situ methods," *Remote Sens.*, vol. 2, pp. 497–513, 2010.
- [5] Y. Deng, B. S. Manjunath, and H. Shin, "Color image segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1999, pp. 446–451.
- [6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [7] M. Turker and E. H. Kok, "Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping," *Int. Soc. Photogramm. Remote Sens. J. Photogramm. Remote Sens.*, vol. 79, pp. 106–121, 2013.
- [8] Y. Xue, J. Zhao, and M. Zhang, "A watershed-segmentation-based improved algorithm for extracting cultivated land boundaries," *Remote Sens.*, vol. 13, 2021, Art. no. 939.
- [9] S. Ghaffarian and M. Turker, "An improved cluster-based snake model for automatic agricultural field boundary extraction from high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 40, pp. 1217–1247, 2019.
- [10] J. Chen, M. Deng, P. F. Xiao, M. H. Yang, and X. M. Mei, "Rough set theory based object-oriented classification of high resolution remotely sensed imagery," *J. Remote Sens.*, vol. 14, pp. 1139–1155, 2010.
- [11] C. Karydas, S. Gewehr, M. Iatrou, G. Iatrou, and S. Mourelatos, "Olive plantation mapping on a sub-tree scale with object-based image analysis of multispectral UAV data; Operational potential in tree stress monitoring," *J. Imag.*, vol. 3, 2017, Art. no. 57.
- [12] L. Lu, Y. Tao, and L. Di, "Object-based plastic-mulched landcover extraction using integrated Sentinel-1 and Sentinel-2 data," *Remote Sens.*, vol. 10, 2018, Art. no. 1820.
- [13] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *Int. Soc. Photogramm. Remote Sens. J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, 2017.
- [14] L. Xu, D. Ming, W. Zhou, H. Bao, Y. Chen, and X. Ling, "Farmland extraction from high spatial resolution remote sensing images based on stratified scale pre-estimation," *Remote Sens.*, vol. 11, 2019, Art. no. 108.
- [15] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
- [19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.
- [23] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*.
- [24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
- [25] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [26] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
- [27] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.
- [28] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *Int. Soc. Photogramm. Remote Sens. J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.
- [29] X. Zheng, H. Sun, X. Lu, and W. Xie, "Rotation-invariant attention network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 4251–4265, Jan. 2022.
- [30] X. Zheng, T. Gong, X. Li, and X. Lu, "Generalized scene classification from small-scale datasets with multitask learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5609311.
- [31] X. Zheng, X. Chen, X. Lu, and B. Sun, "Unsupervised change detection by cross-resolution difference learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5606616.
- [32] X. Zheng, B. Wang, X. Du, and X. Lu, "Mutual attention inception network for remote sensing visual question answering," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5606514.
- [33] X. Zheng, W. Chen, and X. Lu, "Spectral super-resolution of multispectral images using spatial-spectral residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5404114.
- [34] X. Zheng, X. Chen, and X. Lu, "Visible-infrared person re-identification via partially interactive collaboration," *IEEE Trans. Image Process.*, vol. 31, pp. 6951–6963, Jan. 2022.
- [35] X. Li, Y. Zhou, and F. Wang, "Advanced information mining from ocean remote sensing imagery with deep learning," *J. Remote Sens.*, vol. 2022, 2022, Art. no. 9849645.
- [36] Y. Luo, J. Han, Z. Liu, M. Wang, and G. Xia, "An elliptic centerness for object instance segmentation in aerial images," *J. Remote Sens.*, vol. 2022, 2022, Art. no. 9809505.
- [37] M. Li, J. Li, Y. Liu, and F. Liu, "Detail injection-based convolutional auto-encoder for pansharpening," *J. Remote Sens.*, vol. 2022, 2022, Art. no. 0004.
- [38] A. B. Hamida et al., "Deep learning for semantic segmentation of remote sensing images with rich spectral content," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 2569–2572.

- [39] X. Zhang, B. Wu, L. Zhu, F. Tian, and M. Zhang, "Land use mapping in the Three Gorges Reservoir Area based on semantic segmentation deep learning method," 2018, *arXiv:1804.00498*.
- [40] L. Ding, H. Tang, and L. Bruzzone, "LANet: Local attention embedding to improve the semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 426–435, Jan. 2021.
- [41] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanes, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, 2018, Art. no. 1119.
- [42] G. J. Scott, M. R. England, W. A. Starns, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.
- [43] A. Garcia-Pedrero, C. Gonzalo-Martin, and M. Lillo-Saavedra, "A machine learning approach for agricultural parcel delineation through agglomerative segmentation," *Int. J. Remote Sens.*, vol. 38, pp. 1809–1819, 2017.
- [44] Z. Shuangpeng, F. Tao, and H. Hong, "Farmland recognition of high resolution multispectral remote sensing imagery using deep learning semantic segmentation method," in *Proc. Int. Conf. Pattern Recognit. Artif. Intell.*, 2019, pp. 33–40.
- [45] R. Lu, N. Wang, Y. Zhang, Y. Lin, W. Wu, and Z. Shi, "Extraction of agricultural fields via DASFNNet with dual attention mechanism and multi-scale feature fusion in South Xinjiang, China," *Remote Sens.*, vol. 14, 2022, Art. no. 2253.
- [46] S. Wang, G. Sun, B. Zheng, and Y. Du, "A crop image segmentation and extraction algorithm based on Mask RCNN," *Entropy*, vol. 23, 2021, Art. no. 1160.
- [47] D. Zhang et al., "A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution," *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111912.
- [48] W. Liu et al., "Farmland parcel mapping in mountain areas using time-series SAR data and VHR optical images," *Remote Sens.*, vol. 12, 2020, Art. no. 3733.
- [49] A. Taravat, M. P. Wagner, R. Bonifacio, and D. Petit, "Advanced fully convolutional networks for agricultural field boundary detection," *Remote Sens.*, vol. 13, 2021, Art. no. 722.
- [50] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.
- [51] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 173–190.
- [52] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [53] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
- [55] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [56] M. Everingham, S. M. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, pp. 98–136, 2015.
- [57] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.
- [58] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 44–57.
- [59] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.
- [60] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 182–186.
- [61] I. Demir et al., "Deepglobe 2018: A challenge to parse the Earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–181.



Yusen Xie received the B.S. degree in geographic information science from the School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China, in 2021, where he is currently working toward the M.S. degree in cartography and geographic information system.

His research interests include remote sensing image processing, deep learning, and transfer learning.

Shaolan Zheng received the B.E. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2011, and the M.S. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2013. She is currently working in Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou, China. Her research interests include remote sensing image processing and application, protection and utilization of cultivated land resources.

Haiyun Wang received the Ph.D. degree in environment science from Guangzhou Institute of Geochemistry, Chinese Academy of Science, in 2023. She is currently working with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou, China. Her research interests include interpretation of remote sensing images and geographic information system.

Yuzhou Qiu received the B.S. degree in geographic information science from Guangzhou University in 2017. She is currently working with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou, China. Her research interests include semantic segmentation and scene classification for high-spatial resolution remote sensing imagery.

Xilan Lin received B.S. degree in geographic information science from Guangzhou University, in 2019. She is currently working with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou, China. Her research interests include remote sensing image classification and geographic information system.



Qian Shi (Senior Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from the Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2015.

She is currently an Associate Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote sensing image classification, including

deep learning, active learning, and transfer learning.